

Web Page Recommendation System Using Usage Profile

Publication History

Received: 04 August 2015

Accepted: 21 August 2015

Published: 1 September 2015 (Special issue)

Citation

Parkavi C, Rathipriya R. Web Page Recommendation System Using Usage Profile. *Indian Journal of Engineering*, 2015, 12(30), 284-289

Web Page Recommendation System Using Usage Profile

C. Parkavi

Research Scholar

Department of Computer Science
Periyar University, Salem- 636 011
Tamilnadu, India
E-mail: sparkavi11@gmail.com

R. Rathipriya

Assistant Professor

Department of Computer Science
Periyar University, Salem- 636 011
Tamilnadu, India
E-mail: rathi_priyar@yahoo.co.in

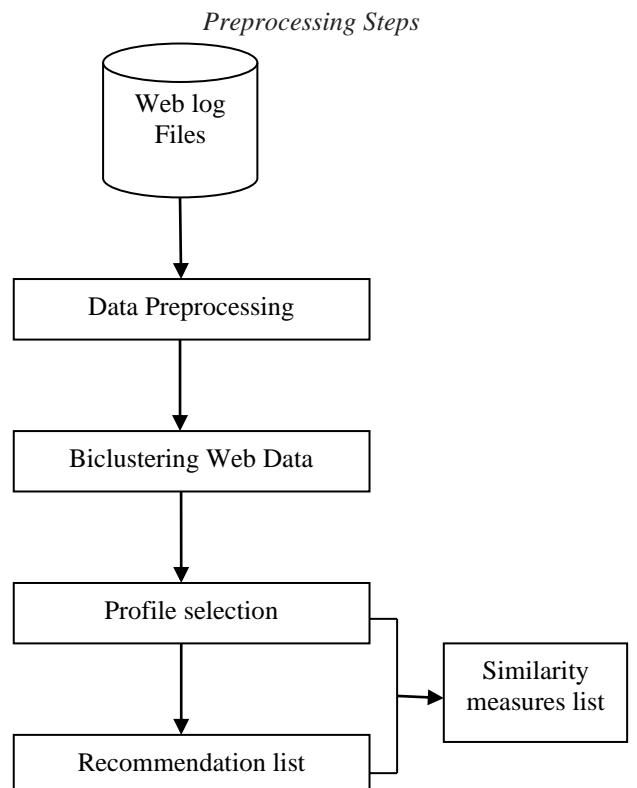
Abstract—This paper focuses on the various similarity measures to increase the recommendation quality in the recommendation system. Biclustering approach is used to the aggregate usage profiles from web usage data. Web recommendation system is proposed using k-means algorithm. It is a simple, classical and efficient algorithm for clustering. Various similarity measures like hamming distance, Jaccard dissimilarity, matching dissimilarity, rogers Tanimoto dissimilarity, Yule dissimilarity, dice similarity, Russell rao dissimilarity, sokalsneath dissimilarity are used to generate the recommended set of pages with high quality. Recommendation process is an online component. Then the standard datasets from the UCI repository have been used to demonstrate the results of the algorithm. The experimental results are encouraging in terms of the quality of the high solutions.

Keywords—recommendationsystem, bicluster, userprofile, similarity measures

I. INTRODUCTION

Web Data Mining is a technique used to crawl through various web resources to collect required information, which enables an individual or a company to promote business, understanding marketing dynamics, new promotions floating on the Internet, etc [1]. Web mining is a new area of research in information technology which is done by using data mining techniques. It is used to discover and extract information from the web data such as content data, structured data, hyperlink data, log data and usage data [2]. Web recommendation system is the subclass of information filtering system. In this web recommendation system has two components are online component and the offline component. A user profile is a visual display of personal data associated with a specific user, or a customized desktop environment. A user profile can also be considered as the computer representation of a user model [3]. A profile can be used to store the description of the characteristics of person. Web usage profile consists of Transaction identification and the page view identification. Transaction identification is implemented the user session and the user access paths are extracted from the web access log information is appended. A page view is an aggregate representation of a collection of web objects contributing to the display on a user's browser

resulting from a single user action (such as a click through). Web usage mining has gained much attention as it is found to fulfill the needs of personalization. In this user profile has to be used as the biclustering method. Biclustering is another data mining technique. It allows simultaneously clustering for rows and columns. Biclustering is simply a general heading of one particular class of data mining techniques. Similarity is an internal relationship between the data objects. A small distance indicates a high degree of similarity and a large distance indicates a low degree of similarity. Similarity measure is subjective and is highly dependent on the domain and application.



II. RELATED WORK

Many researches were given to study the web access based recommendation system in sequential based data mining techniques. To increase the recommendation system has to the similarity measures is an important aspect of this research.

A. Recommender System

In this previous work the web page recommender system has to implement the collaborative filtering using for K nearest neighbor technique is widely used for e-commerce applications. In these users has to the online clustering of user's in the recommendation system. To overcome these limitations, of research has to focus on the web usage mining approach of web personalization [4]. In this research has to aggregate usage profiles can be used in the dynamic recommendation based on the current user's interest.

Most of the recommender system has to implement by using the K-means clustering approach for derive the web usage profile to follow by the recommender system for navigation profile. In this method has to usage profile by using various data mining techniques. Then the techniques have to implement the usage profile of web recommendation system [5].

Web Usage mining approach for web personalization. The pattern discovery phase using various data mining techniques is performed offline to improve the scalability of collaborative filtering. The discovered patterns or aggregate usage profile can be used to provide dynamic recommendations based on the current user's interest.

B. Similarity Measures

The user vector can be obtained by retrieving the access logs of the site. If two users accessed the same pages, they might have some similar interests in the sense that they are interest in the same information (e.g., news, electrical products etc). The number of common pages they accessed can measure this similarity. A new hybrid swarm intelligence based biclustering approach is used for identifying optimal usage profiles based on their browsing interest and provides recommendation on web pages. The proposed approach is tested on msnbc.com dataset. Then the results to indicate the measures have to be modified in the various similarity measures to be used in this paper.

III. RECOMMENDATION SYSTEM BASED ON USAGE PROFILE

The current user is anonymous to the recommender system with no previous navigation history; hence a sliding window technique over the current user session was used to represent the user history. To do so, the user to represent the user current session is broken into two parts; the first part with size n pages is used as the surrogate user history which is matched against the web navigation profiles then produces a recommendation list from the selected profile. The remaining pages from the second part which is used for the comparison purpose to evaluate the recommendation accuracy.

Identify applicable sponsor/s here. If no sponsors, delete this text box (sponsors).

In this section, precision, coverage and F1measure are used to evaluate the recommendation effectiveness (AlMurtadhaet *et al.*, 2010). Let A be active current session taken from the evaluation set and R be a recommendation set generated by using the proposed system over the navigation profiles. W represents the items that already visited by the user in A.

Then the recommendation list has to evaluate by the precision, recall and F1measure in this profiles. Finally, to be calculated in various similarity measures to be used in the best profiles.

For the above, the following similarity measures are used:

1. Hamming distance
2. Matching Dissimilarity
3. Jaccard Dissimilarity
4. Dice Dissimilarity
5. Rogers Tanimoto Dissimilarity
6. Sokal sneath Dissimilarity
7. Yule Dissimilarity

A. Page Weight Calculation

The page weighted calculation to be evaluated by Equ 1

$$\text{Usage profile} = \{p, w(p) \mid p \in P, w(p) \geq \text{min_weight}\} \quad (1)$$

Where

$$P = \{p_1, p_2, \dots, p_n\} \quad (n \text{ is a set of page views})$$

$$W(p) = \text{attribute weights in the bicluster}$$

The user profile selection has to chosen by the maximum weighted value to be chosen in the bicluster. The user profile selection is done by choosing the maximum weighted page in the bicluster. This profile is used to generate the recommended web page set using similarity measures.

B. Precision

Precision is the fraction of retrieved documents that are relevant to the find. In this precision to the ration number of relevant records retrieved to the total number of irrelevant and relevant record retrieved. It is usually expressed as a percentage. Denominator for precision is all that is returned. Equ 2 has to defined by the

$$\text{Precision} (R, A) = \frac{|R \cap (A - W)|}{|R|} \quad (2)$$

R = R is the Recommended pages

A = A is a Active session taken from the evaluation set

W = W represents the items that already visited by the user in A.

Precision is also used with recall, the percent of all relevant documents that is returned by the search. The two measures sometimes used together in the F1 measure or F1 score to provide a single measurement of a system.

C. Recall

The Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. Denominator for recall is all that is relevant. Recall is also called as the coverage.

$$\text{Coverage (R, A)} = \frac{|R \cap (A-W)|}{|(A-W)|} \quad (3)$$

Similarly, R is a recommended page, A is an active session and the W is the Windows size to be calculated by the recall or coverage.

Precision is also referred to as Positive Predictive Value (PPV) other related measures used in classification include true negative rate and accuracy. True negative is also called as the Specificity.

D. F1 Measure

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

There are several reasons that the F score can be criticized in the particular circumstances due to its bias as an evaluation metric.

IV. EXPERIMENTAL RESULTS

The experimental setup of the test performance is coded in MATLAB 6.0 and numerical experiments are performed on PC with Intel core and 2.0 GB memory...

a. Dataset Description

In this section, the performance of the proposed method is evaluated by conducting the datasets from UCI repository datasets. In this datasets has to be taken by the MSNBC dataset for Anonymous Web Data. The data comes from Internet Information Server (IIS) logs for msnbc.com. The web log files of msnbc.com web site have been used to evaluate the performance of the proposed algorithm. The web site includes the page visits of users who visited the msnbc.com web site. Each sequence in the dataset corresponds to page views of a user. The parameters taken for study are number of rows as users and the number of columns as pages and the number of bicluster as population. Then the number of user clusters to be binary conversion of the rows and columns in the bicluster for recommendation system. In this parameter setup has to be discussed as table I.

TABLE I. *PARAMETER SETUP*

Parameters	
No of biclusters	10
No. of Row clusters	2000
No of column clusters	17
Page Weight threshold	0.5

b. Results and Discussion

The usage profile extracted from msnbc dataset is taken for online component for recommendation process. The weight of each page view in the aggregated usage profile is calculated and tabulated. In this result, table II tabulates the list of weighted pages in the bicluster. The user profile has to calculate by the weighted profile in the bicluster. Then the user profile bicluster to perform the all weights of web user navigation. The table III tabulates the list of aggregate usage profiles for the bicluster. The aggregate usage profile is used to evaluate the recommendation quality. The table IV tabulates the recommended list of page using extracted aggregated user profile navigation. Finally, table V tabulates evaluation measures like precision, recall and F1 measure for the proposed work.

TABLE II.

WEIGHTS OF PAGES IN THE BICLUSTER

Profile ID	Pages	Page weights in the profile			
1.	12, 13, 14, 15	0.1887	0.1321	0.0943	0.0566
2.	12, 13, 14, 15	0	0	0	0
3.	12, 13, 14, 15	0.1045	0.0796	0.0547	0.0398
4.	12, 13, 14, 15	0.6265	0.4578	0.2530	0.0964
5.	7, 8	1.0000	0.7547		
6.	7, 8	0.4213	0.2416		
7.	7, 8	0.6219	0.4726		
8.	7, 8	1	1		
9.	1, 2, 3, 4, 5	1	1	1	1
10.	1, 2, 3, 4, 5	1	1	1	1
11.	1, 2, 3, 4, 5	1	1	1	1
12.	1, 2, 3, 4, 5	1	1	1	1
13.	9, 10, 11	0.5472	0.3585	0.3019	
14.	9, 10, 11	0.1067	0.0674	0.0337	
15.	9, 10, 11	0.3433	0.2388	0.1592	
16.	9, 10, 11	1.0000	0.9157	0.7590	
17.	6	1			
18.	6	0.6910			
19.	6	0.8358			
20.	6	1			

In this profile has to implement the biclustering approach of the various cluster method. Then the method to be the weighted values of the page cluster in the biclustering approach. In this profile has to be greater than 5 be the calculated. The user profile has to calculate by the weighted value on $W(P) > \min_weight$ (weighted value greater than 5).

TABLE III. LIST OF AGGREGATE USAGE PROFILE

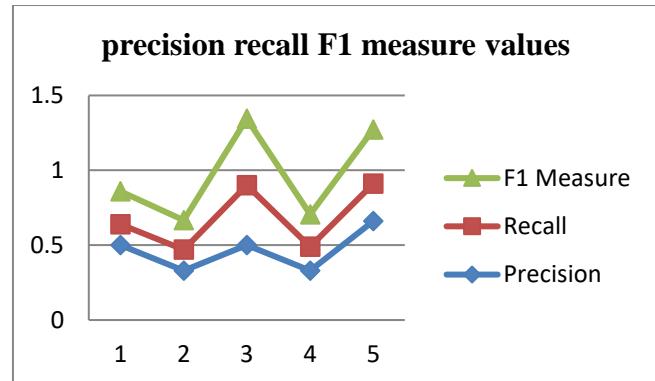
Id	Page Cluster	Profile				
1	12, 13, 14, 15	0.6265	0.4578	0.2530	0.0964	
2	7, 8	1.0000	0.7547			
3	1, 2, 3,4,5	1	1	1	1	1
4	9,10,11	1.0000	0.9157	0.7590		
5	6	0.8358				

TABLE IV. RECOMMENDER LIST

Profile ID	Weights profile	Recommendation list
1	12, 13, 14, 15	3,10
2	7, 8	4,6,5
3	1, 2, 3,4,5	4,7,11,12
4	9,10,11	2,3,10
5	6	4,7,8

TABLE V. EVALUATION MEASURE: PRECISION RECALL AND F1 MEASURE

Profile ID	Precision	Recall	F1 Measure
1	0.5	0.14	0.2187
2	0.33	0.14	0.1965
3	0.5	0.4	0.4444
4	0.33	0.16	0.2155
5	0.66	0.25	0.3626



V. CONCLUSION

Recommendation system has emerged as a powerful tool for helping users to find and evaluate items of interest. It uses a variety of techniques to help users to identify their needs. In this proposed work, a biclustering of web usage data is used to select the maximum visited usage profiles for recommendation process. In-order to increase the recommendation quality, this study is carried out different similarity measure. In future, this work is extended with different validation measures to provide high accuracy recommendation.

References

- [1] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining World Wide Web browsing patterns. *Journal of Knowledge and Information Systems*, (1) 1, 1999.
- [2] R. Cooley, P-T. Tan., and J. Srivastava. WebSIFT: The Web site information filter system. In *Workshop on Web Usage Analysis and User Profiling (WebKDD99)*, San Diego, August 1999.
- [3] Ashwin G. Raiyani, Sheetal S.Pandya, “Dixcovering User Identification Mining Technique for preprocessed web log data”, *Journal of information knowledge and research in computer engineering*, Vol-2, Issue 2, pp:477-482, 2012.
- [4] Delwar H.Arif,AHM Sofi Ullah, KM HAbibullah, Md Ali Al Mamum, “personalized Web Recommendation Combining User-centered Collaborative Technique with URL Weighting”, *International journal of computer applications*, Vol-63, pp:13-18, 2013.
- [5] AlMurtadha, Y. M., & Sulaiman, M. N. B., Mustapha, N. & Udzir, N. I. (2010). Mining web navigation profiles for recommendation system. *Information Technology Journal*, 9(4), 790-796.
- [6] Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2002). Discovery and evaluation of aggregate usage profile for web personalization. *Communication of ACM*, 6(1), 61-82.
- [7] Rathipriya, R., Thangavel, K., & Bagyamani, J. (2011). Evolutionary biclustering of click-stream data. *International Journal of Computer Science Issues*, 8(1), 32-38.
- [8] Rathipriya, R., Thangavel, K., & Bagyamani, J. (2011). Binary particle swarm optimization based Bi-clustering of web usage data. *International Journal of Computer Applications*, 25(2), 43-49.
- [9] K.Suneetha, P.Sunilkumar Reddy, “A Survey on web recommendation systems”, *CSEA2012*, Vol-4, pp: 46-51, 2013.
- [10] Kate A.Smith, Alan Ng, “Web page clustering using a self-organizing map of user navigation patterns”, *Decision Support Systems* 35, pp 245-256, 2003.
- [11] Mobasher, B. (1999). Web Personalizer: A ServerSide Recommender System Based on Web Usage Mining. *Technical Report*, *Telecommunications systems and Information Systems*.

- [12] Manisha Sajwan, Kritika Acharyaa, Sanjay Bhargava, "Swarm Intelligence Based Optimization for web Usage Mining in Recommender System", International journal of computer applications technology and research, Vol-3, Issue-2, pp: 119-124, 2014.
- [13] R. Forsati, M. Meybodi and A. Rahbar, "An Efficient Algorithm for Web Recommendation Systems," AICCSA 2009 IEEE/ACS International Conference on Computer Systems and Applications, Rabat, 10-13 May 2009, pp. 579-586. doi:10.1109/AICCSA.2009.5069385.
- [14] P.S.Raja, R.Rathipriya, "Optimal web page category for web personalization using biclustering approach", International journal of computational intelligence and informatics", Vol-1, pp: 70-74, 2011.